

# A general dynamical statistical model with possible causal interpretation

Daniel Commenges<sup>1,2</sup> and Anne Gégout-Petit <sup>2,3</sup>

February 2, 2008

*1 INSERM, U 875, Bordeaux, F33076, France*

*2 Université Victor Segalen Bordeaux 2, Bordeaux, F33076, France*

*3 IMB, UMR 5251, Talence, F33405, France*

**Summary.** We develop a general dynamical model as a framework for possible causal interpretation. We first state a criterion of local independence in terms of measurability of processes involved in the Doob-Meyer decomposition of stochastic processes, as in Aalen (1987); then we define direct and indirect influence. We propose a definition of causal influence using the concepts of “physical system”. This framework makes it possible to link descriptive and explicative statistical models, and encompasses quantitative processes and events. One of the features of this paper is the clear distinction between the model for the system and the model for the observation. We give a dynamical representation of a conventional joint model for HIV load and CD4 counts. We show its inadequacy to capture causal influences

while on the contrary known mechanisms of HIV infection can be expressed directly through a system of differential equations.

*Keywords:* Causality; causal influence; differential equations; directed graphs; dynamical models; HIV; randomisation; stochastic processes.

## 1 Introduction

The issue of causality has been studied by many philosophers since Aristotle and is of central importance in all branches of science (see Bunge, 1979 and Salmon, 1984). A central question for scientists who use statistics and for statisticians is whether statistical models may help in deciphering causal links. After recognising that correlation is not causation, scientists have tended to use statistical methods as one element among others to help establish causal links. Epidemiologists are particularly cautious, and with good reason, in concluding to causal influences. There has been however a growing interest in developing statistical models able to represent causal influences. From the beginning, graphs have played an important role in representing the set of causal influences. The pioneering work of Wright (1921, 1934) have inspired the more recent developments of structural equation models (Joreskog, 1978) and graphical models (Dawid, 1979; Lauritzen and Wermuth, 1989; Cox and Wermuth, 1996). An approach using the modelling of “potential outcome”, often called the counterfactual approach, has been proposed in the context of clinical trials by Rubin (1974) and further studied by Holland (1986) among others. The counterfactual approach has been extended to the study of longitudinal incomplete data in several papers, the

results of which have been gathered together by van der Laan and Robins (2002). Spirtes, Glymour and Sheines (2000) and Pearl (2000) develop the issue of investigating causality with graphical models.

The counterfactual approach however has been criticised (Dawid, 2000; Geneletti, 2007) and the modelling of potential outcomes raises difficulties when treating truly dynamical problems. In fact another school tackles causality by directly using dynamical models. This approach started in the econometrics literature with Granger (1969) and Schweder (1970) and was more recently developed by several Scandinavian statisticians using the formalism of stochastic processes, and in particular of counting processes (for a review see Eerola, 1994; Aalen and Frigessi, 2007). Of particular interest is the paper by Aalen (1987) which outlines a general approach for defining influences for stochastic processes through the Doob-Meyer decomposition. The most recent developments of the dynamical approach are the method of “dynamic path analysis” of Fosen et al. (2006) and the study of the possibly cyclic directed graphs associated with this definition of influence by Didelez (2007). Defining influence in the stochastic process framework does not ensure that we make relevant causal inference but we believe that it provides a better formalism for tackling this issue than approaches which deal only with random variables.

The aim of this paper is to develop the dynamical approach in a general framework, focusing in particular on causal interpretation, using the concept of system, which was advocated long ago by von Bertalanffy (1968); we attempt to go from the concept of “influence”, which is mathematically defined, to the concept of “causal influence”, which has a physical meaning.

We make a clear distinction between the model for the system and the model for the observations, a classical distinction in automatics (Jazwinsky, 1970) but not in biostatistics. Moreover we link classical epidemiological models and mechanistic models; the latter are not generally taken into consideration in the literature of causal models although (or because) they make explicit use of scientific knowledge.

The paper is organised as follows. In section 2 we develop a criterion of local independence in terms of measurability of processes involved in the Doob-Meyer representation; then we define direct and indirect influence. In section 3 we propose a definition of causal influence using the concepts of “physical system” and “physical laws” for which we propose a definition. Our framework makes it possible to link descriptive and explicative statistical models and encompasses the analysis of events and of quantitative processes. In section 4 we develop the distinction between the model for the system and the model for the observation. In section 5 descriptive and explicative joint models of HIV load and CD4 counts are considered.

## 2 Local independence, direct and indirect influence

### 2.1 Notations

Consider a filtered space  $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$  and a multivariate stochastic process  $\mathbf{X} = (\mathbf{X}_t)_{t \geq 0}$ ;  $\mathbf{X}_t$  takes values in  $\mathfrak{R}^m$ , and the whole process  $\mathbf{X}$  takes values in  $D(\mathfrak{R}^m)$ , the Skorohod space of all cadlag functions:  $\mathfrak{R}_+ \rightarrow \mathfrak{R}^m$ . We suppose

that all the filtrations satisfy the usual conditions. We have  $\mathbf{X} = (X_j, j = 1, \dots, m)$  where  $X_j = (X_{jt})_{t \geq 0}$ . We shall note  $X_j \in \mathbf{X}$ . We denote by  $\mathcal{X}_t$  the history of  $\mathbf{X}$  up to time  $t$ , that is  $\mathcal{X}_t$  is the  $\sigma$ -field  $\sigma(\mathbf{X}_u, 0 \leq u \leq t)$ , and by  $(\mathcal{X}_t) = (\mathcal{X}_t)_{t \geq 0}$  the families of these histories, that is the filtration generated by  $\mathbf{X}$ . Similarly we shall denote by  $\mathcal{X}_{jt}$  and  $(\mathcal{X}_{jt})$  the histories and filtration associated to  $X_j$ . If  $C$  is a subset of  $(1, \dots, m)$  we shall call  $X_C$  the multivariate process  $(X_j, j \in C)$ .

## 2.2 Local independence, direct and indirect influence

Let  $\mathcal{F}_t = \mathcal{H} \vee \mathcal{X}_t$ ;  $\mathcal{H}$  may contain information known at  $t = 0$ , in addition to the initial value of  $\mathbf{X}$ . We shall consider the class of special semi-martingales, that is the class of processes which admit a unique Doob-Meyer decomposition in the  $(\mathcal{F}_t)$  filtration, under probability  $P$ :

$$\mathbf{X}_t = \Lambda_t + M_t, t \geq 0, \quad (1)$$

where  $M_t$  is a martingale and  $\Lambda_t$  is a predictable process with bounded variation. We shall denote the Doob-Meyer decomposition of  $X_j$ :  $X_{jt} = \Lambda_{jt} + M_{jt}$ . We shall consider the non-degenerate case in which all the components of  $M$  are different from zero; the deterministic case will be studied in section 2.4. We shall assume two conditions bearing on the bracket process of the martingale  $M$ :

**A1**  $M_j$  and  $M_k$  are orthogonal martingales, for all  $j \neq k$ ;

**A2**  $X_j$  is either a counting process or is continuous with a deterministic bracket process, for all  $j$ .

We call  $\mathcal{D}$  the class of all special semi-martingales satisfying **A1** and

**A2.** The class of special semi-martingales is stable by change of absolutely continuous probability (Jacod and Shiryaev, 1987, page 43) and this is also true for the the class  $\mathcal{D}$ .

**Definition 1 (Weak conditional local independence (WCLI))**  $X_k$  is weakly locally independent of  $X_j$  in  $\mathbf{X}$  on  $[0, \tau]$  if and only if  $\Lambda_k$  is  $(\mathcal{F}_{-jt})$ -predictable on  $[0, \tau]$ , where  $\mathcal{F}_{-jt} = \mathcal{H} \vee \mathcal{X}_{-jt}$  and  $\mathcal{X}_{-jt} = \bigvee_{l \neq j} \mathcal{X}_{-lt}$ . Equivalently we can say in that case that  $X_k$  has the same Doob-Meyer decomposition in  $(\mathcal{F}_t)$  and in  $(\mathcal{F}_{-jt})$ . We will note in that case  $X_j \not\rightarrow \mathbf{X} X_k$ .

**Remark 1.** Assumption **A2** is necessary for the measurability-based definition of WCLI to be clearly interpreted. If we did not impose **A2** we could find counter-examples in which a WCLI holds while intuitively independence does not hold. Such a counter-example is the process  $\mathbf{X} = (X_1, X_2)$  which is the solution of the differential equation:  $dX_{1t} = a dt + b dW_{1t}$ ;  $dX_{2t} = X_{1t} dt + e^{X_{1t}} dW_{2t}$ , Where  $W_1$  and  $W_2$  are Brownian motions. We would not like to say that  $X_2$  is WCLI of  $X_1$ . However, because  $X_1$  appears in the bracket process of  $X_2$ ,  $\mathcal{X}_{1t}$  is included in  $\mathcal{X}_{2t}$  so that  $\Lambda_2$  is  $\mathcal{X}_{2t}$ -predictable and thus we would conclude that  $X_2$  is WCLI of  $X_1$ .

**Remark 2.** It is tempting to define WCLI directly in terms of the conditional independence:

$$\mathcal{X}_{kt} \perp\!\!\!\perp_{\mathcal{X}_{Ct-}, \mathcal{X}_{kt-}} \mathcal{X}_{jt-}, 0 \leq s < t \leq \tau. \quad (2)$$

Here  $\mathbf{X} = (X_j, X_k, X_C)$ . However, this condition is void in general when we consider processes in continuous time. Because conditional independence is defined via conditional probability and in general, events of  $\mathcal{X}_{kt}$  will have

conditional probabilities equal to one or zero given  $\mathcal{X}_{kt-}$ , the condition will always hold. It is possible that WCLI can be defined in terms of conditional independence of  $\sigma$ -fields but this is an open problem.

**Definition 2 (Direct influence)** *We shall say that if  $X_k$  is not WCLI of  $X_j$  in  $\mathbf{X}$ ,  $X_j$  directly influences  $X_k$  in  $\mathbf{X}$  and we will note  $X_j \longrightarrow_{\mathbf{X}} X_k$ .*

**Definition 3 (WCLI and Direct influence for set of components)** *Let  $A, B$  subsets of  $(1, \dots, m)$ . We shall say that  $X_A \longrightarrow_{\mathbf{X}} X_B$  if there is  $j \in A$  and  $k \in B$  such that  $X_j \longrightarrow_{\mathbf{X}} X_k$ .*

What we call here “direct influence” is the time-continuous analogue of Granger strong causality (Granger, 1969). We may consider another, stronger, condition of local independence.

**Definition 4 (Strong conditional local independence (SCLI))**  *$X_k$  is SCLI of  $X_j$  in  $\mathbf{X}$  if and only if  $X_j \not\rightarrow_{\mathbf{X}} X_k$  and there is no  $X_D \in \mathbf{X}$  such that  $X_j \longrightarrow_{\mathbf{X}} X_D$  and  $X_D \longrightarrow_{\mathbf{X}} X_k$  and we will note  $X_j \not\rightarrow_{\mathbf{X}} X_k$ .*

**Definition 5 (Influence)** *We shall say that if  $X_k$  is not SCLI of  $X_j$ ,  $X_j$  influences (at least indirectly)  $X_k$  in  $\mathbf{X}$  and we will note  $X_j \rightarrow\rightarrow_{\mathbf{X}} X_k$ .*

An interesting case is when weak independence holds but strong independence does not hold; equivalently  $X_j$  influences  $X_k$  but  $X_j$  does not directly influence  $X_k$ : we shall say that  $X_j$  indirectly influences  $X_k$ .

**Definition 6 (Indirect influence)** *If  $X_j \rightarrow\rightarrow_{\mathbf{X}} X_k$  and  $X_j \not\rightarrow_{\mathbf{X}} X_k$  then there is  $X_C \in \mathbf{X}$  such that  $X_j \longrightarrow_{\mathbf{X}} X_C \longrightarrow_{\mathbf{X}} X_k$  and we shall say that  $X_j$  indirectly influences  $X_k$  through  $X_C$  in  $\mathbf{X}$ .*

**Remark.** Since the Doob-Meyer decomposition depends on  $P$  so do all the independencies and influences; realising this fact is crucial for the definition of causal influence in section 3.1.

## 2.3 Differential equation: towards causal interpretation

Writing the process of interest in the form of a stochastic differential equation (SDE) is a way of making the causal mechanisms at work more explicit. If  $\Lambda_t$  is differentiable, the Doob-Meyer decomposition can be written:

$$d\mathbf{X}_t = \lambda_t dt + dM_t, \quad (3)$$

with  $\Lambda_t = \int_0^t \lambda_u du$ . Differential equation models are commonly used in physics, biology and in finance (Oksendal, 2000) to model the evolution of  $\mathbf{X}_t$  as a function of the past plus a random term brought by the martingale. The two main cases, which have been considered in different streams of research, are the case where the trajectories of  $\mathbf{X}$  are continuous and the case where  $\mathbf{X}$  is a counting process. In the case of continuous trajectories of  $\mathbf{X}$  it is common to take for  $M$  a Brownian martingale (in which case  $dM_t = f(t)dW_t$ , with  $W = (W_t)$  a Brownian motion), and the models considered are Itô processes. In the case where  $\mathbf{X}$  is a counting process we write:

$$d\mathbf{X}_t = \lambda_t dt + dM_t, \quad (4)$$

where  $M$  is a discontinuous martingale with predictable variation process equal to  $\Lambda$ , and  $\lambda$  is called the intensity of the process. We may consider mixing the two cases, considering that  $\mathbf{X} = (X_1, X_2)$ , where  $X_1$  is an Itô



process and  $X_2$  a counting process, each of these processes being possibly multivariate. The processes defined by these differential equations are not Markov in general. The Markov assumption is an interesting particular case and it is discussed in section 3.2.2.

## 2.4 The deterministic case

Ordinary differential equation (ODE) models seem to arise as particular cases in which  $M = 0$ . So one way to apply our definition of WCLI to deterministic models is to consider that these models are in fact stochastic but the martingale has a bracket process which takes small values in regard to  $\Lambda$ . Particular phenomena appear in purely deterministic models, in particular because the concept of *filtration* no longer applies. In that case the unicity of the differential equation is lost. For instance consider the process  $\mathbf{X} = (X_1, X_2)$ ; consider the case where the process  $\mathbf{X}$  is deterministic and the trajectories are solutions of the ODE system:  $dX_{1t} = a dt$ ;  $dX_{2t} = X_{1t}dt$  with initial conditions  $X_{10} = X_{20} = 0$ . The trajectories are also solutions of the ODE system:  $dX_{1t} = a dt$ ;  $dX_{2t} = at dt$  with initial conditions  $X_{10} = X_{20} = 0$ . One would be tempted to say that  $X_1$  influences  $X_2$  when looking at the first ODE system and but not when looking at the second one. The second ODE system however is not time-homogeneous. Unicity can thus be restored if we impose the restriction of time-homogeneity (see in section 3.2.2 a discussion of the physical meaning of time-homogeneity). Taking advantage of the unicity of the time-homogeneous differential equation representation, we will consider it as the canonical representation, if it exists. We can then use the definition of WCLI for stochastic differential equations to define WCLI

for the deterministic case: construct a SDE system by adding to the ODE system orthogonal martingales with deterministic brackets. The influence graph of the time-homogeneous ODE system is, by definition, the same as that of the derived SDE. In the above example, if we add a standard Wiener martingale to the canonical (time-homogeneous) representation we obtain the SDE:  $dX_{1t} = a dt + dW_{1t}$ ;  $dX_{2t} = X_{1t}dt + dW_{2t}$ , in which it is clear that we have  $X_1 \longrightarrow_{\mathbf{X}} X_2$ .

## 2.5 Graph representation

We may construct as in Didelez (2007) a directed graph representing influences between components of  $\mathbf{X}$ . This directed graph has for vertices the components  $X_j$  and there is a directed edge  $(j, k)$  if and only if  $X_j \longrightarrow_{\mathbf{X}} X_k$ . Note that there can be two directed edges between two vertices, for instance  $(j, k)$  and  $(k, j)$ ; this can be denoted by two arrows or by a double-sided arrow  $(\longleftrightarrow)$ . A path is an ordered sequence of directed edges  $\{(j_0, j_1), (j_1, j_2), \dots, (j_{k-1}, j_k)\}$ . Indirect influence can be read directly off the graph:  $X_j \rightarrow\rightarrow_{\mathbf{X}} X_k$  if there is a path from  $j$  to  $k$ . An example is shown in figure 1 which represents the hypothetical influence graphs for processes  $\mathbf{X}^1$  on the left and  $\mathbf{X}^2$  on the right; the graphs are not acyclic and in particular we have  $X_2 \longrightarrow_{\mathbf{X}^1} X_4$  and  $X_4 \longrightarrow_{\mathbf{X}^1} X_2$ . We see also that  $X_1$  indirectly influences  $X_4$  but does not influence  $X_3$ , which we can note:  $X_1 \rightarrow\rightarrow_{\mathbf{X}^1} X_4$  and  $X_1 \not\rightarrow_{\mathbf{X}^1} X_3$ . The graph for  $\mathbf{X}^2$  on the right may represent a richer system; we shall develop the issue of considering a family of nested systems in section 3.1.

We say that  $X_C$  blocks the paths from  $X_l$  to  $X_k$  if all the paths from  $X_l$  to  $X_k$  contain a node in  $X_C$ . For instance  $X_4$  blocks the paths from  $X_3$  to

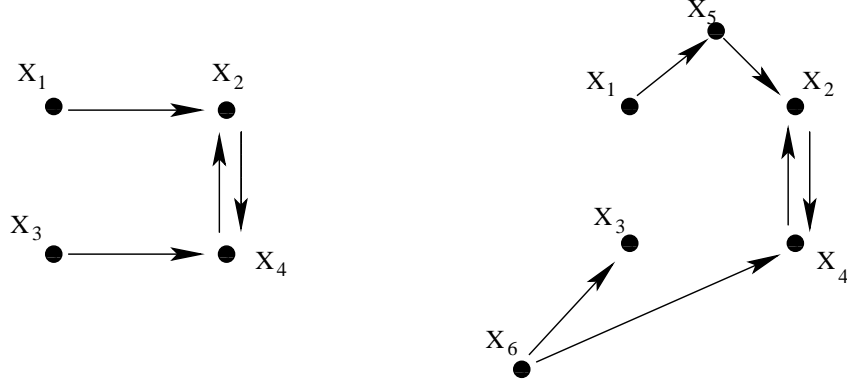


Figure 1: Example of two graphs from the same physical system.

$X_2$  in  $\mathbf{X}^1$ . In  $\mathbf{X}^2$  there is no path from  $X_3$  to  $X_2$ , so  $X_4$  still blocks the paths from  $X_3$  to  $X_2$ , although in a trivial manner. If there is a path from  $X_l$  to  $X_k$  and  $X_j$  blocks the paths from  $X_l$  to  $X_k$  there is necessarily a path from  $X_l$  to  $X_j$  and a path from  $X_j$  to  $X_k$ , which can be expressed as:

**Lemma 1 (Decomposable influence)** *If  $X_l \rightarrow\!\!\rightarrow_{\mathbf{X}^m} X_k$  and  $X_j$  blocks the paths from  $X_l$  to  $X_k$  then  $X_l \rightarrow\!\!\rightarrow_{\mathbf{X}^m} X_j$  and  $X_j \rightarrow\!\!\rightarrow_{\mathbf{X}^m} X_k$ .*

### 3 Causal influences

#### 3.1 Systems, causal influence

In this section we outline a philosophical theory of causality; this theory is necessarily incomplete and questionable but we feel that a theory of this kind is necessary to link the mathematical definitions to the real world. The main concept used will be that of *system* and we will take two examples to illustrate this and other concepts: the first is the archetypal example of the

solar system; the second is the system formed by the immune system and a population of HIV viruses. Thus our first task is to define a system, that we will also call a “physical system”  $\mathcal{S}$  in which we are interested. To define a system we admit that we can define a *level* at which relevant characteristics can be defined: we may distinguish a vector of *attributes* and a *state* vector. The attributes essentially define the system and do not vary in time; the state represents the characteristics, in general varying with time, in which we are really interested and will be represented by a multivariate stochastic process. We may decide for instance that the level we are interested in is that of the sun and the planets and their trajectories. A possible system may be identified by the sun and the nine planets; the attributes of the system are the masses of these ten celestial bodies; the state at time  $t$  is the vector of position and speed (in a reference system) of the ten celestial bodies. What we have excluded in defining this level are the details of the planets such as their physical structure, presence of life, particular events like storms and so on (see Batterman, 2002). In the Immune-HIV system example, we may decide that the level we are interested in is that of populations of cells or of HIV viral particles in a particular subject; the attributes describe which types of cells or viral particles are considered and characteristics of these populations if they may differ from one subject to another; the state may be the numbers (or concentrations) of different types of cells or of viral particles. At this level we are not interested in the fate of a particular cell.

Now we suppose that we are particularly interested in one or several components of the state. We assume that there are laws which govern the evolution of the states and some of the laws tell us that the evolution of

the component  $j$  at time  $t$  depends on the component  $k$  just before  $t$ . Newton's laws, including the gravitation law, tell us how to compute the force of attraction between two massive bodies. However it is impossible to find a system with two massive bodies completely isolated from the rest of the universe; moreover it is very difficult to avoid the circularity in the definitions. For instance we have used the concept of *system* in the previous sentence, a concept which is still not defined. It would be tempting to define *system* by first defining *causal influence*. However in order to define *causal influence* we need to apply natural laws to a system. There is the problem of defining *natural laws* or *physical laws*.

**Definition 7 (System)** *A system is the couple  $\mathcal{S} = (\mathbf{A}, \mathbf{X})$  of attributes and state. The attribute  $\mathbf{A}$  is a possibly random element with value in  $\mathbb{R}^d$  which, together with the state, is sufficient to identify the system. The state  $\mathbf{X}$  is a stochastic process from  $(\Omega, \mathcal{F})$  on  $(D(\mathbb{R}^m), \Sigma)$ , where  $\Omega$  is the universe and  $\mathcal{F}$  contains all the events pertaining to the level of interest;  $D(\mathbb{R}^m)$  is a Skorohod space of all cadlag functions:  $\mathbb{R}_+ \rightarrow \mathbb{R}^m$ , and  $\Sigma$  the Borel sigma-field derived from the Skorohod topology.*

We consider that deterministic  $\mathbf{A}$  and  $\mathbf{X}$  is a particular case of the stochastic case, with the reservation made in section 2.4 that WCLI is defined only for time-homogeneous ODE. Often the attribute will be considered as deterministic. In the solar system example, both attributes and states may be considered as deterministic, or we may consider it as random but work with a probability conditional to the observed value. The rationale for considering attributes as random is that they are the results of systems of another

level: the existence of the planets is the result of the process of formation of the solar system. Note that even in that example, complex systems or long range predictions may raise the issue of *chaos*, thus introducing a stochastic feature (Murray and Dermott, 1999).

Given a system  $\mathcal{S}^m = (\mathbf{A}^m, \mathbf{X}^m)$ , we call  $\mathcal{F}_t^m$  the sigma-field generated by the attribute and the history of the state at time  $t$ ,  $\mathcal{F}_t^m = \sigma(\mathcal{A}^m) \vee \sigma(\mathbf{X}_u^m, 0 \leq u \leq t) = \mathcal{A}^m \vee \mathcal{X}_t^m$ . It is important to consider several systems and in particular we may consider *nested systems*. A system  $\mathcal{S}^{m'}$  is nested in  $\mathcal{S}^m$  if  $\mathcal{F}_t^{m'} \subset \mathcal{F}_t^m$  for all  $t$ :  $\mathcal{S}^{m'}$  can be enlarged by addition of attributes ( $\mathcal{A}^{m'} \subset \mathcal{A}^m$ ) and/or addition of  $\mathbf{X}^{m'}$  components ( $\mathcal{X}_t^{m'} \subset \mathcal{X}_t^m$ ). We can consider a sequence of nested systems  $\mathcal{S} = \{\mathcal{S}^m\}_{m>0}$  (we note  $\mathcal{S}^m \in \mathcal{S}$  and  $\mathcal{S}^m \subset \mathcal{S}^{m'}$  if  $m < m'$ ). In this case, the family  $\{\mathcal{F}_t^m\}_{m>0}$  forms a filtration (for each  $t$ ). If we consider a period of observation  $[0, \tau]$  (included in the definition of the level) we note  $\mathcal{F}^m = \mathcal{F}_\tau^m$ . Note that saying that  $\mathcal{S}^m \subset \mathcal{S}^{m'}$  is more general than considering that all the components of  $\mathcal{S}^m$  belong to  $\mathcal{S}^{m'}$ , although most result will refer to this case.

From now on, we will speak about direct and indirect influences of  $X_j$  on  $X_k$  in the system  $\mathcal{S}^m$  (and denote  $X_j \longrightarrow_{\mathcal{S}^m} X_k$  or  $X_j \rightarrow\rightarrow_{\mathcal{S}^m} X_k$ ) these influences corresponds of the definitions of influences in  $\mathbf{X}^m$  in section 2 with  $\mathcal{H} = \mathcal{A}^m$ .

We assume that there is a true probability law  $P^*$  on  $(\Omega, \mathcal{F})$  and we denote its restriction to  $\mathcal{F}^m$  by  $P_{\mathcal{F}^m}^*$ . We would like to approach  $P_{\mathcal{F}^m}^*$  by applying *physical laws*. We shall now endeavor to define *physical laws*. Let us first define *mathematical laws*.

**Definition 8 (Mathematical laws)** *Mathematical laws at a certain level*

are a set of mathematical procedures that can be applied to any system  $\mathcal{S}^m$  of this level to build a probability  $P^{\mathcal{S}^m}$  on  $\mathcal{F}^m$ .

Generally the probability  $P^{\mathcal{S}^m}$  will be different from  $P_{\mathcal{F}^m}^*$ . Suppose that we are particularly interested in a system  $\mathcal{S}^1$ , we may have to consider richer systems for making correct predictions for the system of interest. We define *physical laws* as yielding a probability that may be as close as we wish from  $P_{\mathcal{F}^1}^*$ , if we can apply them to a correct system.

**Definition 9 (Physical laws)** *If for any system  $\mathcal{S}^1$  of a given level, there exists a sequence of nested systems  $\mathcal{S} = \{\mathcal{S}^m\}_{m>0}$  including  $\mathcal{S}^1$  and mathematical laws such that  $P_{\mathcal{F}^1}^{\mathcal{S}^m}$  converges weakly toward  $P_{\mathcal{F}^1}^*$ , these mathematical laws will be called physical laws at this level, and such a sequence  $\mathcal{S}$  will be called an approximating sequence for  $\mathcal{S}^1$ .*

The weak convergence means that  $\int g dP_{\mathcal{F}^1}^{\mathcal{S}^m} \rightarrow \int g dP_{\mathcal{F}^1}^*$  for any  $\mathcal{F}_1$ -measurable continuous bounded function  $g$  on  $\Omega$ . We may also write  $d_P(P_{\mathcal{F}^1}^{\mathcal{S}^m}, P_{\mathcal{F}^1}^*) \rightarrow 0$ , where  $d_P(.,.)$  is the Prokorov metric for probability measures based on the Skorohod topology. The advantage of the Prokorov metric is that it metrizes weak convergence (Gibbs and Su, 2002) and it encompasses the deterministic case (which makes sense in the solar system example). In the deterministic case  $\mathbf{X}$  takes the value  $X^*$  with probability one under  $P_{\mathcal{F}^1}^*$  and the value  $X^{\mathcal{S}^m}$  under  $P_{\mathcal{F}^1}^{\mathcal{S}^m}$  and we have  $d_P(P_{\mathcal{F}^1}^{\mathcal{S}^m}, P_{\mathcal{F}^1}^*) = d(X^{\mathcal{S}^m}, X^*)$ .

We may postulate the existence of *physical laws*. This postulate reflects the asymptotic *separability* of the universe; that is, for making good predictions we do not need to take into account the whole universe, but on the other hand, application of the laws (even if we know the correct laws) never

leads to perfect prediction, partly because we have isolated a system from the rest of the universe.

The systems may be more or less satisfactory according to the distance to the true probability achieved. For instance we would not call a set constituted of the Earth and Mars a satisfactory system; if we applied Newton's laws to this system we would see that the observed trajectories would be in large disagreement with the predicted ones; we would thus search for a better set of bodies, for instance the set (Sun, Earth, Mars).

We have to make an assumption of finiteness of the approximating sequence to have a clear definition of causal influence. We conjecture that this assumption could be avoided using a quantitative approach of WCLI but this is beyond the scope of this paper.

**A3.** There is a *perfect* system  $\mathcal{S}^M$  for  $\mathcal{S}^1$  such that  $\mathcal{F}^1 \subset \mathcal{F}^M$  and  $P_{\mathcal{F}^1}^{\mathcal{S}^M} = P_{\mathcal{F}^1}^*$ .

This means that the probability law computed with the physical law applied to system  $\mathcal{S}^M$  coincides with the true law on the events of interest  $\mathcal{F}^1$ .

We assume that **A1** and **A2** hold for all the systems considered (see discussion in section 3.2.1); assuming **A3** we can give the following definition.

**Definition 10 (Causal influence)** *A component  $j$  has a causal influence on a component  $k$  in  $\mathcal{S}^1$  if  $X_j \rightarrow\rightarrow_{\mathcal{S}^M} X_k$  under  $P^*$ , if  $\mathcal{S}^M$  is a perfect system for  $\mathcal{S}^1$ .*

**Remark.** The direct influences under the physical law are the same in all the systems and in particular in the perfect system; a direct influence under the physical law is thus always a causal direct influence.



**Example: Solar system.** If we consider a system (Earth, Moon) the law of gravitation tells us that the earth (in our presentation, the position of the Earth) has an influence on the trajectory of the moon; by definition (if we accept that the law of gravitation is a *physical law*), this is a causal influence. Even if this system is not completely satisfactory, the notable fact is that in any richer system, the Earth will have an influence on the Moon; this stability is characteristic of causal influences.

**Example: Immune system-HIV.** The mechanisms which derive from the properties of HIV and CD4 lymphocytes are such that HIV can infect CD4 lymphocytes and that infected lymphocytes can produce viruses. The number of viruses produced depends in part on the number of infected lymphocytes. Thus the component of the state “number of viruses” (in a given individual) has a causal influence on the component “number of infected lymphocytes”. We can deduce the form of causal influences at the level of concentrations from knowledge of the mechanisms which lead to the replication of the virus and application of diffusion laws. The approach is similar to Boltzman’s theory of gases (see Strevens, 2005).

The problem of estimation of the true law  $P^*$  will be dealt with in section 4.

## 3.2 Stability of structures in sequences of systems

### 3.2.1 Stability of the class $\mathcal{D}$

Since influences are defined in the class  $\mathcal{D}$  and there is a need to consider sequences of systems, the stability of  $\mathcal{D}$  in the sequence is crucial. We have

the following Lemma:

**Lemma 2 (Stability of  $\mathcal{D}$ )** *If  $X^M \in \mathcal{D}$ , then  $X^m \in \mathcal{D}$  for all  $X^m \subset X^M$ .*

**Proof.** The class of special semi-martingales is stable by change of filtration (Jacod and Shiryaev, 1987). The optional square-bracket process does not depend on the filtration. Thus it remains deterministic for continuous processes (A2) and the martingales remain orthogonal (A1). For counting processes the orthogonality of the martingales holds if and only if the martingales cannot jump at the same time, which does not depend on the filtration; the martingales of a continuous and a counting process are always orthogonal.

### 3.2.2 Instability of the homogeneous Markov property

It is interesting to examine the Markov properties of the models. In the general case the derivatives of the predictable processes involved in the Doob-Meyer decomposition depend on the whole past of the process. For instance we can make these dependencies explicit by writing  $\lambda(t) = \lambda(t, X_u, 0 \leq u < t)$ . In Markov models these functions depend only on the present, or more precisely on  $X_{t-}$ :  $\lambda(t, X_u, 0 \leq u < t) = \lambda(t, X_{t-})$ . The model is (time)-homogeneous if these functions do not depend on time:  $\lambda(t, X_{t-}) = \lambda(X_{t-})$ . Typical physical models are time-homogeneous Markov models. The Markov property means that knowledge of the past cannot improve our knowledge of the future if we know the present, and the homogeneous property means that the laws of the universe we have used for constructing the model do not

change. So one can argue that if the model we consider is not a homogeneous Markov model we have omitted important components in the model.

If a process is time-homogeneous Markov in  $\mathcal{S}^M$  under  $P^*$  this does not in general hold for  $\mathcal{S}^m \subset \mathcal{S}^M$ ; thus the homogeneous Markov property is not stable in  $\mathcal{D}$ . This fact explains why it is often needed to consider non-homogeneous and even non-Markovian models in biology; indeed the systems considered are often oversimplified in view of the complexity of the real systems, leading to a loss of the time-homogeneous Markov property.

### 3.2.3 Faithfulness and stability of influences

To go further, we assume that  $P^*$  is “faithful”, a property which is discussed for instance in Robins et al. (2003) for directed acyclic graphs, and that we define in our context as:

**Definition 11 (Faithful probability)** *A probability  $P$  is faithful for a sequence  $\mathcal{S}$  if for any  $\mathcal{S}^{m'}, \mathcal{S}^m \in \mathcal{S}$  such that  $\mathcal{S}^{m'} \subset \mathcal{S}^m$  and such that  $X_j^m = X_j^{m'} = X_j$  and  $X_k^m = X_k^{m'} = X_k$ , we have  $X_j \longrightarrow_{\mathcal{S}^m} X_k$  implies  $X_j \longrightarrow_{\mathcal{S}^{m'}} X_k$ . Equivalently,  $X_j \not\longrightarrow_{\mathcal{S}^{m'}} X_k$  implies  $X_j \not\longrightarrow_{\mathcal{S}^m} X_k$ .*

Figure 1 illustrates a case which is compatible with  $P$  being faithful: if  $\mathcal{S}^1$  (resp.  $\mathcal{S}^2$ ) has the left (resp. right) influence graph, we see that the weak independence between  $X_1$  and  $X_3$  is stable when the system is enriched from  $\mathcal{S}^1$  to  $\mathcal{S}^2$ ; on the other hand the influence of  $X_3$  on  $X_4$  disappears ( $X_6$  acts as a confounder process); finally the direct influence of  $X_1$  on  $X_2$  becomes an indirect influence through  $X_5$ . Faithfulness does not hold in general; however one may argue that it does not hold only in very specific cases. We show for instance in Appendix A that

**Proposition 1 (faithful-diffusion)** *If the system  $\mathcal{S}^M = (\mathbf{A}^M, \mathbf{X}^M)$  is such that  $\mathcal{A}^M = \{\emptyset, \Omega\}$  and  $\mathbf{X}^M$  is a linear time-homogeneous diffusion process under  $P$ , then faithfulness holds for any sequence of nested systems  $\mathcal{S} = (\mathcal{S}^1, \dots, \mathcal{S}^M)$ , where  $\mathcal{S}^m = (\mathbf{A}^m, \mathbf{X}^m)$  is a system such that  $\mathbf{X}^m \in \mathbf{X}^M$ .*

Even in the true probability the influences for  $\mathbf{X}$  may be non-causal. However, with the faithfulness assumption two conclusions can be drawn: (i) if  $X_1 \rightarrow\rightarrow_{\mathcal{S}^{m'}} X_2$  then either this influence is causal or one can find a  $\mathcal{S}^m$ ,  $\mathcal{S}^{m'} \subset \mathcal{S}^m \in \mathcal{S}$ , in which there is a process which influences both  $X_1$  and  $X_2$  (a common ancestor in graph terminology): such a process may be called a confounder in epidemiological terminology; (ii) if  $X_1 \not\rightarrow\rightarrow_{\mathcal{S}^{m'}} X_2$  this means that  $X_1$  does not have a causal influence on  $X_2$ . If an indirect influence in  $\mathcal{S}^{m'}$  is causal it is stable by considering richer systems  $\mathcal{S}^m$ ; direct influences in  $\mathcal{S}^{m'}$  may be related to indirect causal influences in  $\mathcal{S}^m$ .

Now we study criteria of independence of processes, which leads us to a mathematical proof in our context of the causal interpretation of a direct influence of a randomized process (our Theorem 1).

**Definition 12 (Dynamical independence)** *If  $X_j \not\rightarrow\rightarrow_{\mathcal{S}^m} X_k$ ,  $X_k \not\rightarrow\rightarrow_{\mathcal{S}^m} X_j$  and  $X_j$  and  $X_k$  have no common ancestor, we say that  $X_j$  and  $X_k$  are dynamically independent in  $\mathcal{S}^m$ .*

**Definition 13 (Non-influenced process)** *A process  $X_A \in \mathbf{X}^m$  is a non-influenced process in  $\mathcal{S}^m = (\mathbf{A}^m, \mathbf{X}^m)$  for probability  $P$  if  $X_j \not\rightarrow_{\mathcal{S}^m} X_A$  for  $P$ , for all  $X_j \in \mathbf{X}^m$ .*

**Lemma 3 (Component and group Dynamical independence)** *If  $X_j$  and  $X_k$  are dynamically independent in  $\mathcal{S}^m = (\mathbf{A}^m, \mathbf{X}^m)$ , it is possible to find*

$X_A, X_B, X_C$  such that  $\mathbf{X}^m = (X_A, X_B, X_C)$  where  $X_A$  and  $X_B$  are non-influenced in  $\mathcal{S}^m$ . Conversely any component, say  $X_j$ , of  $X_A$  and any component of  $X_B$ , say  $X_k$ , are dynamically independent in  $\mathcal{S}^m$ .

If  $X_C$  is influenced by both  $X_A$  and  $X_B$ , the influence graph of  $\mathbf{X}^m = (X_A, X_B, X_C)$  where  $X_A$  and  $X_B$  are non-influenced is  $X_A \longrightarrow X_C \longleftarrow X_B$

**Lemma 4 (Dynamical independence and independence)** *Let  $\mathbf{X}^m = (X_A, X_B, X_C)$ . Consider the assumptions : (a)  $X_C \not\rightarrow_{\mathcal{S}^m} X_A$  and  $X_C \not\rightarrow_{\mathcal{S}^m} X_B$ , (b)  $P$  is faithful for the sequence  $\mathcal{S} = (\mathcal{S}^2, \mathcal{S}^m)$  with  $\mathcal{S}^2 = (\mathbf{A}, \mathbf{X}^2)$  with  $\mathbf{X}^2 = (X_A, X_B)$  and (c) the decomposition of  $(X_A, X_B)$  in  $\mathcal{S}^m$  is that of a diffusion with jumps such that given  $\mathcal{A}$  the corresponding SDE satisfies the uniqueness conditions in law.*

*Consider the two following propositions :*

- (i)  $X_A$  and  $X_B$  are independent conditionally on  $\mathcal{A}$ ;
- (ii)  $X_B \not\rightarrow_{\mathcal{S}^m} X_A$ ,  $X_A \not\rightarrow_{\mathcal{S}^m} X_B$  and  $X_{A0} \perp_{\mathcal{A}} X_{B0}$ .

*Then, under assumptions (a) and (b), (i) implies (ii). Moreover if (c) holds, the converse is true.*

**Remark.** Diffusion with jumps and conditions of uniqueness are given in Jacod and Shiryaev (III.2). In proposition (ii)  $X_{A0}$  and  $X_{B0}$  are the initial values of  $X_A$  and  $X_B$ .

Although the Lemma may seem intuitively obvious a general proof is not simple to find. See Appendix B for an outline of proof.

**Definition 14 ( $\mathcal{S}$ -Non-influenced process)** *Let a system  $\mathcal{S}^1$  belonging to a sequence  $\mathcal{S}$ ; a process  $I \in \mathbf{X}^1$  is a  $\mathcal{S}$ -non-influenced process for probability  $P$  if whatever  $\mathcal{S}^m \subset \mathcal{S}$ ,  $I \perp\!\!\!\perp \mathcal{A}^m$  and  $X_j^m \not\rightarrow_{\mathcal{S}^m} I$  for  $P$ , for all  $j$ .*

The only clearly non-influenced processes for  $P^*$  are randomised processes, generally randomised attribution of a treatment. In observational studies, the non-influenced quality will always be an assumption. For instance a genetic factor may in some circumstances be considered as non-influenced (Didelez and Sheenan, 2007). However, in our approach genetic factors would generally be considered as (observed or non-observed) attributes, and not part of the state.

**Theorem 1 (Non-influence and causality)** *Let  $\mathcal{S}$  an approximating sequence for  $\mathcal{S}^1 = (\mathbf{A}^1, (I, X_j))$ . Suppose that  $P^*$  is faithful for any sequence in the associated perfect system  $\mathcal{S}^M$ , that  $(I, X_j)$  satisfies the assumption (a) and (c) of Lemma 4 and  $I_0 \perp\!\!\!\perp_{\mathcal{A}^m} X_{j0}$  for all  $m$ . If  $I$  is a  $\mathcal{S}$ -non-influenced process for  $P^*$  and  $I \rightarrow_{\mathcal{S}^1} X_j$  for  $P^*$ , then  $I$  causally influences  $X_j$ .*

**Proof.** If  $I$  did not causally influence  $X_j$ , we would have  $I \not\rightarrow_{\mathcal{S}^M} X_j$  for  $P^*$ . Since  $I$  is a non-influenced process, according to Lemma 3 and Lemma 4,  $I \perp\!\!\!\perp_{\mathcal{A}^M} X_j$  and using the fact that  $I \perp\!\!\!\perp \mathcal{A}^m$  for all  $m$  it implies  $I \perp\!\!\!\perp_{\mathcal{A}^1} X_j$ , and in particular that  $I \not\rightarrow_{\mathcal{S}^1} X_j$ , in contradiction with our assumption. Hence the Theorem.

It is interesting to give a version of the idea of instrumental variables (Stock, 2001; Angrist et al., 1996; Greenland, 2000) applied to our context; here the idea is applied only to assess the causal nature of an influence, while

it is often used to estimate the magnitude of the causal influence in specific models. We have the following result:

**Lemma 5 (Instrumental processes)** *Under the assumptions of Theorem 1, if  $I$  is a  $\mathcal{S}$ -non-influenced process,  $I \longrightarrow_{\mathcal{S}^1} X_k$ , and  $X_j$  blocks the paths from  $I$  to  $X_k$  in system  $\mathcal{S}^M$ , then  $X_j$  causally influences  $X_k$ .*

**Proof.** By Theorem 1 we have  $I \longrightarrow_{\mathcal{S}^1} X_k \implies I \rightarrow\rightarrow_{\mathcal{S}^M} X_k$ . If  $X_j$  blocks the paths from  $I$  to  $X_k$  in  $\mathbf{X}^M$ , then by Lemma 1 we have  $X_j \rightarrow\rightarrow_{\mathcal{S}^M} X_k$ ; hence the Lemma.

### 3.3 Implications for physics, system biology and epidemiology

#### 3.3.1 Physics

Let us consider as an example the level of the trajectories of planets in the solar system. The physical system is the set of planets simplified to points in three-dimensional space and we are interested only in their trajectories. The state of the system can be represented by a multivariate process  $\mathbf{X}$ , the components of which are the positions and the speeds of the planets in a given set of axes and this process obeys a differential equation of the type  $d\mathbf{X}_t = g(\mathbf{X}_t)dt$ , where  $g(\cdot)$  is a function derived from Newton's law of mass attraction.  $\mathbf{X}$  is a time-homogeneous Markov process, although degenerated because deterministic. There do not seem to be processes that can be manipulated in this system. However we believe that the influence of a planet on the trajectory of another planet may be considered as being of causal nature.

A first instance of the application of physical laws is to predict or to control the state of the system: for instance one can predict eclipses or control the trajectory of a space vessel. In this case we assume that we know the physical law and that we have a good system.

A second instance is that there is a discrepancy between  $P^*$  and  $P^S$  for the chosen system. If there is not much doubt about the physical laws we are applying (here Newton's laws) then it may be deduced that the system considered is not satisfactory and that it must be increased. A famous example of such an instance is the discrepancy which appeared between the computed and the observed trajectories of Uranus. Leverrier made computations which lead to the discovery of Neptune in 1846. He assumed that the discrepancy in the observed trajectory of Uranus with respect to what was computed using Newton's laws was due to the presence of another planet: he gave the computed position of this planet to Johann Galle and Louis D'Arrest who found it.

A third instance occurs if in spite of refining the system, a discrepancy persists. Then the physical laws may be cast into doubt.

### **3.3.2 Systems biology**

The model is constructed with partially known mechanisms but some of the influences are unknown and even when causal influences are assumed, their precise forms are unknown. These models can be used to test whether some causal influences exist or to quantify them when they are assumed to exist. We will develop the analysis of the interaction between HIV and the immune system in section 5.



### 3.3.3 Epidemiology

Most epidemiological studies endeavour to test the influence of a single factor on a disease process. The physical system contains all biological phenomena implied in the disease as well as the factor of interest; in general there is no physical law, only biological plausibility of some causal influences. A typical system is  $\mathbf{X} = (F, D, C)$  where  $F$  is the factor of interest,  $D$  represents the disease and  $C$  are other processes taken from the system. Such a problem is most often modelled with random variables rather than with stochastic processes. The stochastic process framework allows to take into account the dynamics of the phenomena: typically  $D$  would be a counting process and the exposure factor  $F$  may also vary in time, as is most often the case in reality. The interest often lies in the possible causal influence of  $F$  on  $D$ . Testing whether  $F \longrightarrow_{\mathbf{X}} D$  is generally expressed by saying that we test whether  $F$  is a risk factor for  $D$  by an analysis adjusted on  $C$ . It should be possible to formalize in our framework the condition of “no unmeasured confounders” which makes it possible to conclude that  $F$  causally influences  $D$ . This however requires further work.

In many simple clinical trials the main interest lies in a particular influence, that of a drug on a clinical endpoint. The aim is to test whether there is a causal influence without trying to understand which basic causal mechanisms may explain it, even if there is a biological plausibility that a certain molecule (or treatment) may have a causal influence on the clinical endpoint considered. That is, most often, we do not have physical laws. This is why randomised trials have been developed. If  $F$  is a treatment that can be randomised, it becomes  $\mathcal{S}$ -non-influenced. Then by Theorem 1 it is suffi-

cient to look at the influence of  $F$  on  $D$  in any model to deduce the presence or absence of causal influence.

## 4 Model for the observations

In most applications we do not have precise physical laws. Instead of a unique probability we use a model, that is a family of probability  $(P_\theta^{S^m})_{\theta \in \Theta}$  on  $\mathcal{F}^m$ . The choice of the model may include scientific knowledge, that is a model can be considered as an incompletely specified physical law. If the system  $\mathcal{S}^m$  is rich enough (ideally if it is the "perfect" system  $\mathcal{S}^M$ ) and if the knowledge incorporated in the model is correct, the model is well-specified, that is  $P_{\mathcal{F}^m}^* \in (P_\theta^{S^m})_{\theta \in \Theta}$ . Even if the model is not well specified it is interesting to find the value  $\theta_0$  such that  $P_{\theta_0}^{S^m}$  is the closest to  $P_{\mathcal{F}^m}^*$ . Since the latter is unknown we need observations, which by definition are realisations of  $\mathcal{F}^m$ -measurable random variables under probability  $P_{\mathcal{F}^m}^*$ . Generally complex systems will be observed with complex observations schemes, leading to incomplete (or coarsened) or indirect observations. Generalising the approach of Heitjan and Rubin (1991) to stochastic processes we may say that the observation, represented by the sigma-field  $\mathcal{O}^m$ , are generated by  $g(\mathbf{X}, G)$ , where  $G$  is a component which may be deterministic or stochastic. If  $G$  is deterministic we have  $\mathcal{O}^m \subset \mathcal{F}^m$ ; if however  $G$  is random,  $\mathcal{O}^m$  is not a subset of  $\mathcal{F}^m$ .

To choose a probability in the model close to  $P_{\mathcal{F}^m}^*$  we must construct an estimator  $\hat{\theta}(\mathcal{O}^m)$ . For maximum likelihood or maximum penalised likelihood estimators we must compute the likelihood for the observation, which is the Radon-Nykodim derivative of  $P_\theta^{S^m}$  relative to a reference probability  $P_0$  on

the sigma-field  $\mathcal{O}^m$ , and we denote it  $\mathcal{L}_{\mathcal{O}^m}^{PS^m}$ . If the mechanism leading to incomplete data (m.l.i.d.) is deterministic this is equal to  $E_{P_0}(\mathcal{L}_{\mathcal{F}^m}^{PS^m}|\mathcal{O}^m)$  and this is relatively easy to compute. If not, the issue of ignorability of the m.l.i.d. arises: if the m.l.i.d. is ignorable we can proceed as if it was deterministic and obtain nevertheless the correct inference. For instance Commenges and Gégout-Petit (2007) computed the likelihood for counting processes observed with a complex, but ignorable, observation scheme. If the m.l.i.d. is not ignorable we have to include  $G$  in the system and consider  $\mathbf{X}^{m'} = (\mathbf{X}^m, G)$ ; we have then by definition  $\mathcal{O}^{m'} \subset \mathcal{F}^{m'}$  and we can apply the above formula. The price to be paid is that we need additional assumptions and the computation of the likelihood may not be easy.

In epidemiology one generally has a sample of observations for a sample of systems indexed by  $i$ ,  $i = 1, \dots, n$ . The most common framework is that the observations are independently identically distributed. In this framework, if we can describe the system and its observation for a generic item, we can do it for the sample; this is why in this paper we always omit the subscript  $i$ .

## 5 Dynamical models for HIV/AIDS

### 5.1 The problematic of AIDS through dynamic influence graphs

AIDS was identified in 1981 as a life-threatening disease due to acquired immunodeficiency. It was found that this immunodeficiency was essentially due to a decrease of the number of CD4+ T-lymphocytes. In 1983 it was

found that this decrease was mainly due to the destructive replication of a virus in CD4+ lymphocytes and this virus was denominated HIV. Thus we can formulate the causal pathway: “presence of HIV causes low CD4 counts which causes AIDS which causes death”. Although most researchers would agree with this phrase and think that what is behind the word “cause” are particular biological mechanisms which could be further reduced to biochemical laws, it remains vague because i) time is only implicitly involved through the fact that cause precedes effect; ii) each modality is relative to another modality (presence vs absence, low vs high and so on).

The dynamical model approach allows us to make the causal statement more precise. First we construct the processes  $I = (I_t)$ ,  $T = (T_t)$ ,  $A = (A_t)$ ,  $D = (D_t)$ :  $I$  is a counting process representing HIV infection,  $T$  has a continuous state-space and represents CD4+ T-lymphocytes count;  $A$  and  $D$  are counting processes representing AIDS and death respectively. We can express the causal structure by the influence graph:

$$I \longrightarrow T \longrightarrow A \longrightarrow D$$

Indeed we know from the results of research (involving virology, immunology, and clinical research) that these influences can be interpreted as causal. It is interesting to note that we consider that  $I \longrightarrow T$  is causal although it is difficult to manipulate  $I$ . A more detailed description of the infection can be made by introducing the viral load process  $V = (V_t)$ . There is of course a direct influence of  $I$  on  $V$  because if  $I_t = 0$  then  $dV_t = 0$ . When considering the evolution of infected subjects the process of interest is  $V$  (not  $I$  which is identically equal to one in these patients).

## 5.2 From descriptive to mechanistic models

In the conventional epidemiological and biostatistical literature, linear mixed-effect models have been used to analyse separately repeated measurements of CD4 counts and viral load. For instance to analyse viral load following initiation of a highly active anti-retroviral therapy (HAART) a linear-mixed effect model with two slopes has been used (Jacqmin-Gadda et al., 2000). Potential observations  $Y_j$  are the viral load at time  $t_j$ , or a logarithmic transformation of the viral load; for simplicity we will ignore these normalising transformations here. Some data may be missing (a non-ignorable mechanism here):  $Y_j$  was observed only if  $Y_j > \eta$ , where  $\eta$  is a detection limit, while  $1_{Y_j > \eta}$  was always observed. The model can be written as:

$$Y_j = \beta_0 + a_0 + (\beta_1 + a_1 + \gamma_1 A) \min(t_j, t_*) + (\beta_2 + a_2 + \gamma_2 A)(t_j - t_*) I_{t_j > t_*} + \varepsilon_j, \quad (5)$$

where  $\beta_0, \beta_1, \beta_2$ , are parameters for the intercept, first and second slopes respectively and  $a_0, a_1, a_2$ , are independent normal random effects on the intercept, first and second slopes respectively;  $t_*$  is the time of change of slope (supposed known),  $A$  indicates the treatment and  $\varepsilon_j$  are normal variables with zero expectations: they may be independent or have a correlation structure. In the dynamical model representation, this model can be written in terms of the process  $V = (V_t)$  living in continuous time, representing the concentration of virus at time  $t$ . There are at least two ways of representing the random effects: they could be degenerate components of the state or they could be random attributes. We adopt the latter which leads to the simplest expression:

$$dV_t = [(\beta'_1 + \gamma_1 A) I_{t \leq t_*} + (\beta'_2 + \gamma_2 A) I_{t > t_*}] dt + \sigma dW_t, \text{ with } Z_0 = \beta'_0 \quad (6)$$

where  $\beta'_0$  is a random initial condition and  $\beta'_1$  and  $\beta'_2$  are considered as random attributes; the link with the above model is that  $\beta'_j$  has expectation  $\beta_j$  and variance  $\text{var } a_j$ . The observation (treating the observation times as fixed) is  $\mathcal{O} = \sigma(1_{Y_j > \eta}, 1_{Y_j > \eta} Y_j, j = 1, \dots, m)$  where  $Y_j = V_{t_j} + \varepsilon'_j$ . Note that the error  $\varepsilon_j$  of model (5) is the sum of the value of the martingale at  $t_j$  and the observation error in model (6): we have  $\varepsilon_j = W_{t_j} + \varepsilon'_j$ . The models for the observations may be the same if the correlation structure of the  $\varepsilon_j$  in model (5) is compatible with that produced by model (6). The graph of this process is not very interesting since only  $A$  influences  $Z$ .

A more elaborate model was proposed by Thiébaud et al. (2005). This was a multivariate linear mixed model for jointly modelling viral load and CD4, together with a possibly informative drop-out. For each of the two markers there were two slopes with a fixed and a random effect (as in the previous model). We leave aside here a certain number of features of that paper, including modelling of the drop-out and of explanatory variables, to focus on how the link between observations of HIV load and CD4 counts was modelled. The model can be written:

$$Y_j^1 = \beta_0^1 + a_0^1 + (\beta_1^1 + a_1^1) \min(t_j, t_*) + (\beta_2^1 + a_2^1)(t_j - t_*)I_{t_j > t_*} + \varepsilon_j^1,$$

$$Y_j^2 = \beta_0^2 + a_0^2 + (\beta_1^2 + a_1^2) \min(t_j, t_*) + (\beta_2^2 + a_2^2)(t_j - t_*)I_{t_j > t_*} + \varepsilon_j^2.$$

where  $\varepsilon_j^1$  and  $\varepsilon_j^2$  are zero expectation normal variables. For fixed  $j$ ,  $\varepsilon_j^1$  and  $\varepsilon_j^2$  are independent; the sequences  $\varepsilon_j^k, j = 1, \dots, m$  for  $k = 1, 2$  may be formed of independent variables or have a correlation structure. The link between HIV load and CD4 counts was expressed by correlations of the random effects  $a_l^1$  and  $a_l^2$ ,  $l = 0, 1, 2$ . In particular we could expect negative correlations

between the slopes of HIV load and CD4 counts, which was indeed observed when fitting the model to the data of a therapeutic trial (better viral response was correlated to better immune response).

The model can be expressed in the dynamical framework as:

$$dV_t = [\beta_1'^1 I_{t \leq t_*} + \beta_2'^1 I_{t > t_*}]dt + \sigma_1 dW_{1t}, \text{ with } V_0 = \beta_0'^1$$

$$d\bar{T}_t = [\beta_1'^2 I_{t \leq t_*} + \beta_2'^2 I_{t > t_*}]dt + \sigma_2 dW_{2t}, \text{ with } \bar{T}_0 = \beta_0'^2$$

where  $V_t$  is the logarithm of the viral load and  $\bar{T}_t$  the CD4 counts at time  $t$ ,  $\beta_l'^k = \beta_l^k + a_l^k, k = 1, 2; l = 1, 2$ . As in the previous model there are several ways of treating the random effects; for instance we may consider them as random attributes. The observation is

$$\mathcal{O} = \sigma(1_{Y_{j>\eta}^1}, 1_{Y_{j>\eta}^1} Y_j^1, Y_j^2, j = 1, \dots, m) \text{ with } Y_j^1 = V_{t_j} + \varepsilon_j'^1 ; Y_j^2 = \bar{T}_{t_j} + \varepsilon_j'^2 \quad (7)$$

It is clear from the differential equations above that there is no influence of  $V$  on  $\bar{T}$  whatever the values of the parameter: the influence graph is made of two disconnected vertices. We might have treated the random effects as ancestors, but in this representation too, there is no direct nor indirect influence of  $V$  on  $\bar{T}$ . In this model  $\bar{T}$  is SCLI from  $V$  which does not fit with the known mechanism of the infection. So although this model succeeded in fitting the data better than separate linear mixed models, it is unable to capture any relevant causal influence.

There are different models in which we can express that viral load influences CD4. Having made a clear distinction between the “model for the system” and the “model for the observation” it is natural to construct a model including components that are not observed at all, but which will be

more satisfying with respect to the way it represents the biological mechanisms. One may distinguish infected and un-infected cells and take into account the causal influences in the ODE system (Ho et al., 1995; Perelson et al., 1996). Still a more satisfying model distinguishes between quiescent ( $Q$ ) and activated CD4 ( $T$ ) and between infectious ( $V_I$ ) and non-infectious ( $V_{NI}$ ) virus. Note that distinguishing quiescent and activated CD4 is a way of enriching the state without simply adding a new component. To write the differential equation for the model one uses additional assumptions which are plausible in view of the knowledge of the biological mechanisms: for instance we assume that new CD4+ T lymphocytes are produced (by the thymus) at a rate  $\lambda$ , that only activated cells can be infected, that the probability of meeting of a cell and a virion is proportional to the product of their concentrations. The model proposed by Guedj, Thiébaut and Commenges (2007) was:

$$\begin{aligned}
dQ_t &= (\lambda + \rho T_t - \alpha Q_t - \mu_Q Q_t)dt \\
dT_t &= (\alpha Q_t - (1 - \eta 1_{\{I_t^{RT}=1\}})\gamma T_t V_{It} - \rho T_t - \mu_T T_t)dt \\
dT_t^* &= [(1 - \eta 1_{\{I_t^{RT}=1\}})\gamma T_t V_{It} - \mu_{T^*} T_t^*]dt \\
dV_{It} &= (\omega \mu_{T_t^*} \pi T_t^* - \mu_v V_{It})dt \\
dV_{NI} &= [(1 - \omega) \mu_{T_t^*} \pi T_t^* - \mu_v V_{NI}]dt
\end{aligned}$$

where  $I^{RT}$  is the process indicating whether a treatment based on an inhibitor of the reverse transcriptase is taken at time  $t$ . If we consider the framework of a controlled clinical trial this process is non-influenced and



controlled (because its trajectory is obtained by randomisation). Guedj, Thiébaut and Commenges (2007) assumed that some parameters were random. Such parameters may be considered as random attributes while fixed parameters may be considered as constants of a “physical law”. Note that the system is time-homogeneous, which is satisfactory from an explanatory point of view. Moreover, as we noted in section 2.4, this makes it possible to draw the influence graph of a deterministic model. We could also consider a stochastic differential equation system but inference in this context is very challenging. The observation is the same as in (7), with  $V_{t_j} = V_{It_j} + V_{NI_{t_j}}$  and  $\bar{T}_{t_j} = Q_{t_j} + T_{t_j} + T_{t_j}^*$ .

We could consider mixing this model for the markers with a model for an event such as an opportunistic disease, adding the component  $D = (D_t)$  which is a counting process. The risk of the opportunistic disease may be considered as depending on the concentration of CD4+ T lymphocytes, so that keeping the framework of a time-homogeneous Markov model we can propose a proportional hazard model (but with constant base-line risk  $\gamma$ ):

$$dD_t = I_{\{D_t=0\}} \gamma \exp(\beta_1 Q_t + \beta_2 T_t + \beta_3 Z) + dM_t,$$

where  $Z$  is an explanatory variable. The graph for such a model is given in Figure 2.

Note that if the treatment was an inhibitor of protease the graph would be different: the inhibitor of protease influences  $V_I$  and  $V_{NI}$ . Also, in an observational study, the treatment is in fact influenced by the information on the clinical and biological state of the patient. If we want to represent this situation we have to include the medical doctor in the system: the doctor may decide to modify the treatment after having been informed of

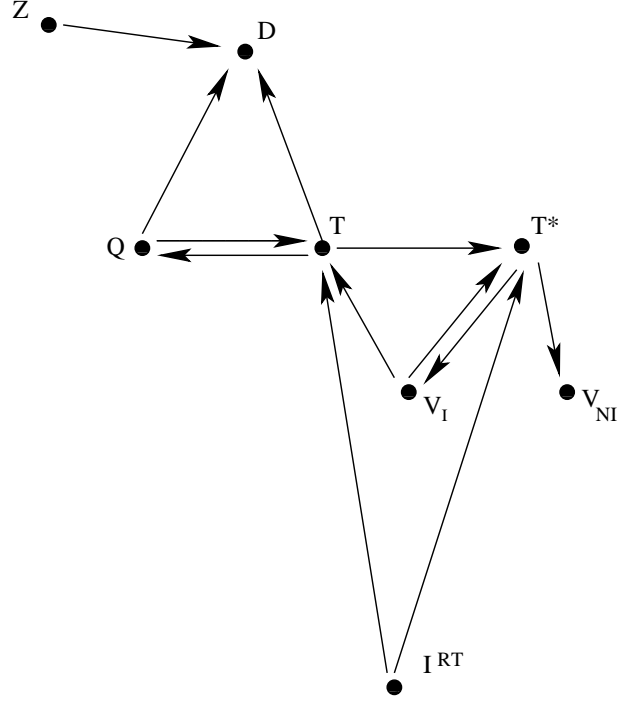


Figure 2: Graph for the mechanistic HIV model.

the measurement of viral load (VL) and of CD4 counts (CD4); note that the processes VL and CD4 are different from  $V$  and  $\bar{T}$  because they carry the information on measurements of these processes, that is  $(V_t, \bar{T}_t)$  carry the observation contained in  $\mathcal{O}$  (see (7)) up to time  $t$ . Then the graph could be as shown in Figure 3, where we have represented by dotted lines the influences of the marker processes on their measurements and the influence of these measurements on the treatment, through the decision of the doctor.

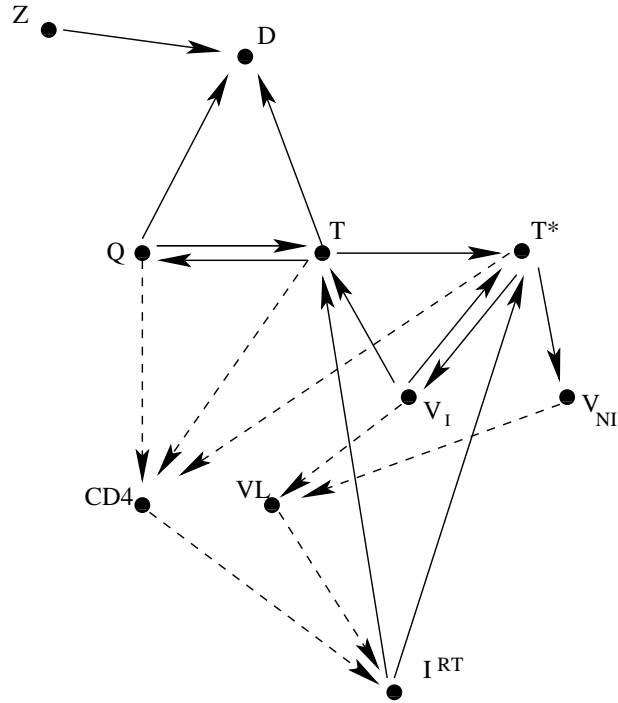


Figure 3: Graph for the mechanistic HIV model.

## 6 References

- Aalen, O. (1987) Dynamic modelling and causality. *Scand. Actuarial J.*, **1987**, 177-190.
- Aalen, O. and Frigessi, A. (2007) What can statistics contribute to a causal understanding? *Scand. J. Statist.*, **34**, 155-168.
- Angrist, J., Imbens, G., and Rubin, D. (1996) Identification of causal effects using instrumental variables. *J. Am. Statist. Ass.*, **91**, 444-455.
- Batterman, R.W. (2002) *The devil in the details*. Oxford University Press.

- Bunge, M. (1979) *Causality and modern science*. Dover Publications, New-York.
- Commenges, D. and Gégout-Petit, A. (2007) Likelihood for generally coarsened observations from multi-state or counting process models. *Scand. J. Statist.*, **34**, 432-450.
- Commenges, D., Joly, P., Gégout-Petit, A. and Liqueur, B. (2007). Choice between semi-parametric estimators of Markov and non-Markov multi-state models from generally coarsened observations. *Scand. J. of Statist.*, **34**, 33-52.
- Cox, D. and Wermuth, N. (1996) *Multivariate Dependencies: Models, Analysis and Interpretation*. Chapman & Hall, London.
- Dawid, A. P. (1979) Conditional independence in statistical theory (with Discussion). *J. Roy. Statist. Soc. A*, **41**, 1-31.
- Dawid, A. P. (2000) Causal inference without counterfactuals. *J. Am. Statist. Ass.*, **95**, 407-448.
- Didelez, V. (2007). Graphical models for composable finite Markov processes, *Scand. J. Statist.*, **34**, 169-185.
- Didelez, V. and Sheehan, N.A.(2007) Mendelian randomisation as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, in press.
- Eerola, M. (1994). *Probabilistic Causality in Longitudinal Studies*. Lecture Notes in Statistics, Springer, New-York.

- Fosen, J., Ferkingstad, E., Borgan, O and Aalen, O.O. (2006) Dynamic path analysis: a new approach to analyzing time-dependent covariates. *Lifetime Data Analysis*, **12**, 143-167.
- Geneletti, S. (2007) Identifying direct and indirect effects in a non-counterfactual framework. *J. Roy. Statist. Soc. A*, **69**, 199-215.
- Gibbs, A.L. and Su F.E. On Choosing and Bounding Probability Metrics. *Int Stat. Rev.* **70**, 419-435.
- Gill, R. D., van der Laan, M. J. and Robins, J.M. (1997). Coarsening at random: characterizations, conjectures and counter-examples, in: *State of the Art in Survival Analysis*, D.-Y. Lin and T.R. Fleming (eds), Springer Lecture Notes in Statistics 123, 255-294.
- Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424-438.
- Greenland, S. (2000) An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.*, **29**, 722-729.
- Guedj, J., Thiébaut, R. and Commenges, D. (2007) Maximum Likelihood Estimation in Dynamical Models of HIV. *Biometrics*, in press.
- Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M. and Markowitz, M.(1995) Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 1995, 373(6510), 123-126.
- Holland, P. (1986) Statistics and Causal Influence. *J. Am. Statist. Ass.*, **81**, 945-960.

- Jacqmin-Gadda, H., Thiébaud, R., Chêne, G. and Commenges, D. (2000) Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics*, **1**, 355-368.
- Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems fo Stochastic Processes*. Berlin: Springer-Verlag.
- Jazwinsky, A.H. (1970) *Stochastic processes and filtering theory*, Academic Press, New-York.
- Joreskog, K. (1978) Structural analysis of covariance and correlation matrices. *Psychometrika* **43**, 443-477.
- Murray, C.D. and Dermott, S.F. (1999) *Solar System Dynamics*. Cambridge University Press.
- Oksendal, B. (2000) *Stochastic Differential Equations*. Springer Verlag, New-York.
- Pardoux, E., (1991) Filtrage non linéaire et équations aux dérivées partielles stochastiques associées. *École d'Été de Probabilités de Saint-Flour XIX—1989*, Lecture Notes in Math., **1464**, 67-163, Springer, Berlin.
- Pearl, J. (2000) *Causality : Models, Reasoning, and Inference*, Cambridge University Press.
- Perelson, A.S., Neuman, A.U., Markowitch, M., Leonard, J.M. and Ho, D.D. (1996) HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science*, **271**, 1582-1586.
- Revuz, D. and Yor, M. (1991) *Continuous Martingales and Brownian Motion*,

New-York: Springer-Verlag.

Robins, J., Scheines, R., Spirtes, P. and Wasserman, L. (2003) Uniform consistency in causal inference. *Biometrika*, **90**, 491, 515

Rubin, D. B., (1974) Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, **66**, 688- 701.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.

Spirtes, P., Glymour, C. and Scheines, R. (2000) *Causation, Prediction, and Search*, MIT Press, Cambridge, Massachussetts.

Strevens, M. (2005) How Are the Sciences of Complex Systems Possible? *Philosophy of Science*, **72**, 531-556.

Schweder, T. (1970) Composable Markov processes. *J. Appl. Probab.*, **7**, 400-410.

Thiébaud, R., Jacqmin-Gadda, H., Babiker, A. and Commenges, D. (2005) Joint Modelling of bivariate longitudinal data with informative drop-out and left-censoring, with application to the evolution of CD4+ cell count and HIV RNA viral load in response to treatment of HIV infection. *Statist. Med.*, **24**, 65-82.

van der Laan, M. and Robins, J. (2002) *Unified methods for censored longitudinal data and causality*, Springer, New-York.

von Bertalanffy, L. (1968). *General System Theory*. New-York: George

Braziller.

Wright, S. (1921) Correlation and causation. *Journal of Agricultural Research*, **20**, 557-585.

Wright, S. (1934) The method of path coefficients. *Ann. Math. Statist.*, **5**, 161-215.

## Appendix A: Faithfulness in diffusion processes

We study the faithfulness property in the case of a system of linear diffusions. For sake of simplicity we consider a process  $\mathbf{X}^3$ ,  $\mathbf{X}_t^3 = (X_{1t}, X_{2t}, X_{3t})$  where  $X_1, X_2, X_3$  are univariate processes. Let us define the processes  $\mathbf{X}^3$  by the following linear stochastic differential equations with constant coefficients:

$$\begin{cases} dX_{1t} &= (a_1X_{1t} + b_1X_{2t} + c_1X_{3t})dt + dW_{1t} \\ dX_{2t} &= (a_2X_{1t} + b_2X_{2t} + c_2X_{3t})dt + dW_{2t} \\ dX_{3t} &= (a_3X_{1t} + b_3X_{2t} + c_3X_{3t})dt + dW_{3t} \end{cases} \quad (8)$$

with initial conditions  $X_{10} = X_{20} = X_{30} = 0$  and  $(W_1, W_2, W_3)$  are independent Brownian motions. We are interested in the semi-martingale decomposition of  $\mathbf{X}_t^2 = (X_{1t}, X_{2t})$  in its own filtration  $(\mathcal{X}_{1t} \vee \mathcal{X}_{2t})$ . If we note  $E[X_{3t} | \mathcal{X}_{1t} \vee \mathcal{X}_{2t}] = \hat{X}_{3t}$  and using the innovation theorem, we find that:

$$\begin{cases} dX_{1t} &= (a_1X_{1t} + b_1X_{2t} + c_1\hat{X}_{3t})dt + \underbrace{dW_{1t} + c_1(X_{3t} - \hat{X}_{3t})dt}_{=dM_{1t}} \\ dX_{2t} &= (a_2X_{1t} + b_2X_{2t} + c_2\hat{X}_{3t})dt + \underbrace{dW_{2t} + c_2(X_{3t} - \hat{X}_{3t})dt}_{=dM_{2t}} \end{cases} \quad (9)$$

$M_{1t}$  and  $M_{2t}$  are independent Brownian motions in the filtration  $(\mathcal{X}_{1t} \vee \mathcal{X}_{2t})$ .



If we suppose that the coefficient  $b_1 \neq 0$ , the probability would not be faithful, if  $X_2 \not\rightarrow_{\mathbf{X}^2} X_1$  that is  $b_1 X_{2t} + c_1 \hat{X}_{3t} = f(\mathcal{X}_{1t})$ . We use the linear filtering equations given in (Pardoux, 1991):

$$\begin{aligned} d\hat{X}_{3t} = & \left[ X_{1t}(a_3 - R_t(a_1 c_1 + a_2 c_2)) + X_{2t}(b_3 - R_t(b_1 c_1 + b_2 c_2)) + \hat{X}_{3t}(c_3 - R_t(c_1^2 + c_2^2)) \right] dt \\ & + R_t c_1 dX_{1t} + R_t c_2 dX_{2t} \end{aligned} \quad (10)$$

$$dR_t = (2c_3 R_t + 1 - R_t^2(c_1^2 + c_2^2))dt \quad (11)$$

A necessary condition in order to delete the dependence of  $X^1$  towards  $X^2$  is that the part directed by  $dW_2$  in  $b_1 X_{2t} + c_1 \hat{X}_{3t}$  equals 0, that is  $(b_1 = -R_t c_1 c_2)$ . If we remark that  $R_t$  which is a solution of the Riccati differential equation (11), cannot be constant, we conclude that the model is faithful.

Now if we suppose that the coefficients are no longer constant and are deterministic time functions and if we suppose that the following relation is true

$$b_1(t) = -R_t c_1(t) c_2(t) \quad (12)$$

The part driven by  $dX_{2t}$  in  $(b_1(t) X_{2t} + c_1(t) \hat{X}_{3t})$  disappears. According to (12) and noting  $Z_t = b_1(t) X_{2t} + c_1(t) \hat{X}_{3t}$ , (for convenience, we sometimes omit the dependence of the coefficients  $(b_1(t), b_2(t), b_3(t), c_1(t), c_2(t), c_3(t))$  on  $t$ )

$$\begin{aligned} dZ_t &= b_1(t) dX_{2t} + c_1(t) d\hat{X}_{3t} + (b'_1(t) X_{2t} + c'_1(t) \hat{X}_{3t}) dt \\ &= c_1(t) \left[ X_{1t}(a_3 - R_t(a_1 c_1 + a_2 c_2)) + X_{2t}(b_3 - R_t(b_1 c_1 + b_2 c_2)) + \hat{X}_{3t}(c_3 - R_t(c_1^2 + c_2^2)) \right] dt \\ &\quad + R_t c_1 dX_{1t} + \left[ (b'_1(t) X_{2t} + c'_1(t) \hat{X}_{3t}) \right] dt \\ &= X_{1t} c_1(t) (a_3 - R_t(a_1 c_1 + a_2 c_2)) + X_{2t} [c_1(t) (b_3 - R_t(b_1 c_1 + b_2 c_2)) + b'_1(t)] \\ &\quad + \hat{X}_{3t} [c_1(t) (c_3 - R_t(c_1^2 + c_2^2)) + c'_1(t)] dt + R_t c_1 dX_{1t} \end{aligned}$$

$Z_t$  is the solution of a stochastic differential equation only driven by  $X_{1t}$  if :

$$c_1(t) [c_1(t)(b_3 - R_t(b_1c_1 + b_2c_2)) + b'_1(t)] = b_1(t) [c_1(t)(c_3 - R_t(c_1^2 + c_2^2)) + c'_1(t)]$$

Using (12) to substitute  $b_1(t)$  we can show that if  $b_3 = R_t(b_2c_2 + c'_2 + c_2 - c_2c_3) + R_{2t}c_2^3$ ,  $Z_t$  is driven by  $X_{1t}$  and the property of faithfulness falls.

This case is extreme. In fact if it holds, the dynamic of  $b_1(t)$ ,  $b_2(t)$  and  $b_3(t)$  is imposed by those of  $c_1(t)$ ,  $c_2(t)$  and  $c_3(t)$

## Appendix B: Proof of Lemma 4

### Proof:

Let us first prove that (i) implies (ii). Consider the Doob-Meyer decomposition of  $X_A$  in the filtration  $\mathcal{A} \vee \mathcal{X}_{At}$ :  $X_{At} = \Lambda_{At} + M_{At}$ . By (i), we have  $E[M_{At} - M_{As} | \mathcal{A} \vee \mathcal{X}_{As} \vee \mathcal{X}_{Bs}] = E[M_{At} - M_{As} | \mathcal{A} \vee \mathcal{X}_{As}]$  and thus the Doob-Meyer decomposition of  $X_A$  is the same in the filtrations  $(\mathcal{A} \vee \mathcal{X}_{At})$  and  $(\mathcal{A} \vee \mathcal{X}_{At} \vee \mathcal{X}_{Bt})$ . This implies  $X_B \not\rightarrow \mathcal{S}^2 X_A$ . By symmetry, we have  $X_A \not\rightarrow \mathcal{S}^2 X_B$  and (ii) follows in  $\mathcal{S}^2$ . Now by the faithfulness property, we have (ii) in all system  $\mathcal{S}^m$  with  $\mathcal{S}^2 \subset \mathcal{S}^m$ .

As for the converse, we prove it in the case of a process satisfying a SDE governed by a Brownian motion in  $(\mathcal{F}_t)$  with  $\mathcal{F}_t = \mathcal{A} \vee \mathcal{X}_{At} \vee \mathcal{X}_{Bt}$ :

$$X_{At} = X_{A0} + \int_0^t f(\mathcal{X}_{As}, \alpha_0) ds + \int_0^t \sigma_{As} dW_{As} \quad (13)$$

$$X_{Bt} = X_{B0} + \int_0^t g(\mathcal{X}_{Bs}, \beta_0) ds + \int_0^t \sigma_{Bs} dW_{Bs} \quad (14)$$

where  $(\alpha_0, \beta_0)$  is  $\mathcal{A}$ -measurable,  $\sigma_A$  and  $\sigma_B$  are deterministic **(A2)** and  $W_A$  and  $W_B$  are two independent Brownian motions **(A1)**. We suppose that

given  $(\alpha_0, \beta_0) \in \mathcal{A}$ , the SDE satisfies assumption assuring uniqueness in law (see for instance Revuz and Yor, 1991: Definitions IX.1.3 and IX.1.4 and Corollary IX.1.14 for the conditions). As by assumption,  $X_A$  and  $X_B$  are non-influenced in  $(\mathcal{A}, \mathbf{X}^m)$ , then whatever the system  $\mathcal{S}^{m'} = (\mathbf{A}, \mathbf{X}^{m'})$  such as  $\mathbf{X}^2 \subseteq \mathbf{X}^{m'} \subseteq \mathbf{X}^m$  the process  $(X^A, X^B)$  always satisfies the same SDE.

However, we can take a new probability space  $(\Omega', \mathcal{F}')$  endowed with two independent Brownian motions  $W_{A'}$  and  $W_{B'}$  and construct two independent processes  $X_{A't} - X_{A'0}$  and  $X_{B't} - X_{B'0}$  on it with  $(X_{A'0}, X_{B'0}) =^{\mathcal{L}} (X_{A0}, X_{B0})$  and with  $X^{A'}$  satisfying SDE (13) driven by  $W_{A't}$  and  $X^{B'}$  satisfying SDE (14) driven by  $W_{B't}$ . By the first part of this demonstration, the decomposition of  $(X_{A'}, X_{B'})$  in  $\mathcal{X}_{A'} \vee \mathcal{X}_{B'}$  is given by the joint system of the 2 equations satisfied by  $X_{A'}$  and  $X_{B'}$  in her own filtration. The vector  $(X_A, X_B)$  and  $(X_{A'}, X_{B'})$  satisfies the same SDE, by uniqueness in law this implies the conditional independence between  $X_A$  and  $X_B$  given  $\mathcal{A}$ .

Using the same reasoning one can extend the result to any diffusion system with jumps (defined in Jacod and Shiryaev, 1987: p. 155) satisfying the condition of uniqueness in law given  $\mathcal{A}$  (see theorem III.2.32 in Jacod and Shiryaev, 1987).